Real-time People Counting Method with Surveillance Cameras Implemented on Embedded System

Hsiang-Chieh Chen, Ya-Ching Chang, Nai-Jen Li, Cheng-Feng Weng, and Wen-June Wang

Abstract—This work presents an image-based method to estimate the number of people crowding in a large-spaced indoor environment. The method is implemented on an embedded system using a digital signal processor. Moving people in each video frame are first extracted using background subtraction method. Additionally, an adaptation scheme is employed to update background model for conditions of sudden lighting changes and abandoned objects. The extracted foreground objects are further accumulated as weighted pixels to estimate the number of people, where the weights are automatically calculated by considering camera settings. Experimental results demonstrate that the proposed method performs well on estimating the number of people in the video surveillance.

Index Terms—Digital signal processor, people counting, video surveillance

I. INTRODUCTION

VISUAL surveillance systems with image processing techniques have recently developed in many applications on embedded-based platforms due to the advancement of technology in industries. Generally a surveillance system could be designed as a distributed computing structure using a server and some digital signal processors (DSPs) for various applications. Besides, a DSP also provides the capability of complicated computation to replace a PC in specific applications. In [1], the data of images are first extracted by a PC and thus transmitted to DSPs for further processing. An intelligent lighting control system implemented on embedded platform with novel image processing algorithms is introduced in [2]. The position and number of occupants analyzed using a DSP are employed to control indoor lights. A face detection system based on a DSP is presented [3]. The research introduced in [4] not only performs a recognition algorithm but also proposes some schemes including data packaging and loop optimization to improve the speed of program. A moving object detection method is realized on a DSP called TMS320DM643 and its corresponding code is particularly rewritten for better performance [5].

This work proposes a novel people counting method for estimating the number of people crowding in large spaces such as lobbies, wholesale- and super-stores. Thus the estimated results could be used in different applicable scenarios, such as lighting and air conditioning control, for advancement of our lives or better efficiency of energy use. This paper is organized as follows. Section II introduces a weighting model related to various monitoring distances and an adjustable human template for counting people from extracted foreground objects. Section III describes the main procedure of our algorithm. Finally, the experimental results and conclusions are respectively presented in Sections IV and V.

II. WEIGHTING MODEL AND HUMAN TEMPLATE

In this work, the number of people present is mainly calculated from the accumulated pixels of foreground objects. Unfortunately, objects in each image capture probably have optical distortion. Fig.1 illustrates the condition that an identical person standing at different positions, A’ and B’, ahead of a camera forms different image sizes. Thus the weighting model presented below is employed to regulate those formed image sizes to identical results while the foreground pixels are accumulated by weightings.

A. Weighting Model

Fig. 2 shows a lateral view of the proposed method in which θ and φV respectively denote the angle of depression and a half of vertical view angle. The camera is set at point H. Fig. 3 illustrates the relationship between the capture frame of surveillance video and real world. The trapezoidal area ABCD is monitored and captured as an image EFCD. In other words, the area ABCD is a real ground region and is projected on the screen of image frame EFCD on the green line RFI in Fig. 2. First, a half of vertical view angle φV is calculated using (1).

\[
\phi^V = \tan^{-1}(H_{ccd} / 2f_{ccd}),
\]

where \( H_{ccd} \) is the height of CCD device in a camera and \( f \) is the focal length of used lens. Assume the captured image is sized by 320 pixels x 240 pixels that is frequently used in surveillance systems and refer to Fig. 2 again. The following equations could be simply derived.

\[
\overline{Hf} = \overline{HI}/\cos(90^\circ - \theta - \phi^V),
\]
Fig. 1. People at different positions result in different sizes in an image.

Fig. 2. The lateral view of setting up a camera.

Fig. 3. The relationship between the image frame and real world.

Fig. 4. The length in real world at $y = 239$ and $y = 240$.

Fig. 5. Top view illustrating the real width at $y = 239$ and $y = 240$.

Finally, the weighting model is presented in the following equation:

$$\text{weight}(y) = \frac{W_y}{W_{240}} \cdot y = 1, 2, ..., 240. \quad (12)$$

B. Human Template

An adjustable human template that presents a virtual person standing at $y = 240$ in proposed to estimate the number of people. The human template is plotted as $\overline{IJ}$ in Fig. 6, in which $\overline{IJ} = \overline{HI} \times \tan(90^\circ - \theta - \phi^v)$. \((13)\)

and $\overline{HI} = \overline{HI} \times \cos(90^\circ - \theta - \phi^v)$. \((14)\)

By applying similar triangles between $\Delta HIJ$ and $\Delta UV$, the $\overline{UV}$ is derived as $\overline{UV} = \overline{IV} \times \sin(90^\circ - \theta - \phi^v)$. \((15)\)

Then we get $\overline{UV}$ that $\overline{UV}$ projects on $\overline{OJ}$ using (16).

$$\overline{JY} = \overline{IY} - \overline{OH} \times \tan(\phi^v - \alpha), \quad (16)$$

where $\alpha = \tan^{-1}\left[\overline{UV} / (\overline{HI} - \overline{HI} \times \cos(90^\circ - \theta - \phi^v))\right]$. \((17)\)
III. MAIN ALGORITHM

The main object of this work is to estimate the number of people using the above-mentioned weighting model and human templates. Additionally the proposed algorithm is implemented on embedded system with a DSP. The main procedure of our algorithm is described in the following four subsections.

A. Image pre-processing

The image pre-processing provides useful features from captured images. A background model is constructed using the first captured image for the background subtraction method [6] to extract foreground objects. Morphological processing including erosion and dilation operations is employed to eliminate foreground noises.

B. Background Updating Scheme

A robust background model used to extract a desired foreground is critical. In this work, a simple and widely-used method [7] is utilized to update the background model to achieve a good balance between performance and computational complexity. The method first saves the m-th and 2m-th image frames from sequence captures. By subtracting these two images pixel by pixel, the corresponding pixels whose results are smaller than a predefined threshold have to be updated. Fig. 8 plots the flowchart of background updating scheme, where the predefined threshold is set as 10).

C. Image Down-sampling for DSP

The format of video source is YUV 4:2:2 and captured frame is sized by 640 × 480 pixels. A downsizing scheme is necessary since the image processing takes a great deal of computation of a DSP. Fig. 9 demonstrates the down-sampling method that takes an average value of four pixels. For example, we take the average of $U_{0,1}$, $U_{0,2}$, $U_{1,1}$ and $U_{1,2}$ to be the down-sized pixel $U_{0,1}$.

D. People Number Estimation

Foreground objects are extracted by subtraction method and then be processed by sequence of erosion and dilation operations. Next the proposed weighting model is utilized to calculate each pixel of foreground to obtain the accumulative weighted pixels using (21).

$$\sum \theta \times \text{weight}(y).$$  \hspace{1cm} (21)

where $N_{\text{foreground}} = \sum_{y=0}^{240} N(y) \times \text{weight}(y)$. (21)

Thus the number of people $N_{\text{people}}$ is calculated

$$N_{\text{people}} = N_{\text{foreground}} / N_{\text{sample}}.$$  \hspace{1cm} (22)

The main steps of the proposed method are summarized below.
Step 1: Capture frame from a surveillance camera.
Step 2: Downsizing the captured frame.
Step 3: Construct background model from the first frame.
Step 4: Update background model.
Step 5: Extract foreground using background subtraction.
Step 6: Construct the weighting model and human templates.
Step 7: Accumulate the weighed pixels.
Step 8: Determine the number of people.

IV. EXPERIMENTAL RESULTS

The proposed method was implemented on an embedded system with a DSP called TMS320DM6437 provided by Texas Instruments. Performance of counting people was evaluated using the videos from large spaces such as lobbies and wholesale stores. Situations with different crowd densities were tested to observe the accuracy of people counting, as shown in Figs. 10(a) and 11(a). In the first experiment, the setup height of camera was 316 cm from floor and the angle of depression was 18.6°. In this experiment, 200 image frames were captured as the frame rate was set as 1 frame per second. Fig. 10(b) plotted the counting results, where the blue line and red one respectively denoted the real and estimated number of people. The mean absolute error (MAE) calculated from those frames was 0.45 people per frame. Similarly, Fig. 11(b) demonstrated the second experiment whose MAE was 0.92 people per frame. Either in the first experiment with fewer people or in the second experiment with more people, the counting MAE in these cases was less than 1 people per frame. The experimental results tested in this work evidently demonstrated the well performance and accuracy of people counting.

V. CONCLUSIONS

This work addresses a method for estimating the number of people. Additionally, the method is further implemented on an embedded system. To deal with optical deformation, a weighting model related to positions ahead of cameras is first introduced and is applied to calculate weighted foreground pixels. Consequently, human templates are employed to count people from the extracted foreground objects or groups. From the experiments with different situations, the proposed method has good performance on estimating the number of occupants.

REFERENCES